Netflix Siblings

Taylor Blair

1/26/2021

#A brief intro

My sister and I have a long history of watching Netflix.

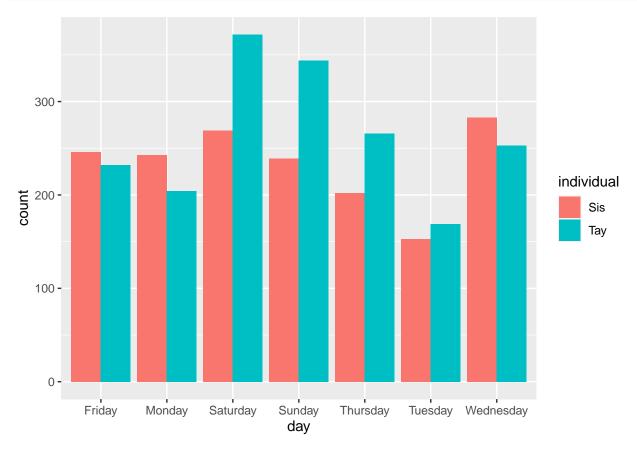
There are some minor outliers in the datasets, namely when we use each others accounts.

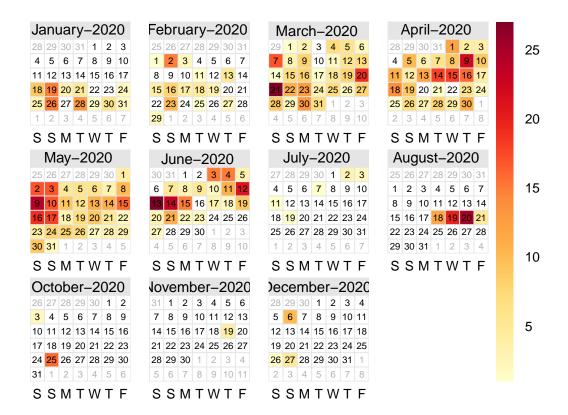
#Loading data and libraries

```
library(tidyverse)
## -- Attaching packages ------ tidyverse 1.3.0 --
## v ggplot2 3.3.3 v purrr
                              0.3.4
## v tibble 3.0.5 v dplyr 1.0.3
## v tidyr 1.1.2 v stringr 1.4.0
## v readr
          1.4.0 v forcats 0.5.0
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()
                   masks stats::lag()
library(openair)
library(readr)
my_history <- read_csv("MyNetflixHistory.csv")</pre>
## -- Column specification -----
    Title = col_character(),
##
    Date = col_character()
sibling_history <- read_csv("SisNetflixHistory.csv")</pre>
##
## -- Column specification -----
    Title = col_character(),
    Date = col_character()
## )
#Data prep
##Date fixing
my_history$Date <- as.Date(my_history$Date,"%m/%d/%y")</pre>
sibling_history$Date <- as.Date(sibling_history$Date,"%m/%d/%y")
```

```
##Movie or TV series
my_history$type <- grepl(":", my_history$Title)</pre>
my_history$type <- my_history$type %>%
  replace(my_history$type==TRUE, "TV Series") %>%
  replace(my_history$type==FALSE, "Movie")
sibling_history$type <- grepl(":", sibling_history$Title)</pre>
sibling history$type <- sibling history$type %>%
  replace(sibling_history$type==TRUE, "TV Series") %>%
  replace(sibling_history$type==FALSE, "Movie")
##Series, Season and episode
my_history <- my_history %>%
  separate(Title,
           c("Series", "Season", "Episode"),
           ": ")
## Warning: Expected 3 pieces. Additional pieces discarded in 267 rows [23, 99,
## 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115,
## 116, 117, ...].
## Warning: Expected 3 pieces. Missing pieces filled with `NA` in 40 rows [5, 50,
## 56, 87, 88, 151, 152, 153, 291, 300, 301, 306, 308, 337, 379, 501, 561, 656,
## 806, 901, ...].
sibling_history <- sibling_history %>%
  separate(Title,
           c("Series", "Season", "Episode"),
           ": ")
## Warning: Expected 3 pieces. Additional pieces discarded in 263 rows [105, 106,
## 107, 108, 115, 116, 127, 128, 138, 139, 355, 399, 400, 401, 402, 403, 404, 405,
## 406, 407, ...].
## Warning: Expected 3 pieces. Missing pieces filled with `NA` in 53 rows [1, 110,
## 174, 175, 261, 262, 263, 275, 276, 281, 282, 283, 284, 285, 454, 460, 515, 516,
## 521, 703, ...].
#Day of Week
day_of_week <- function(x){</pre>
 return (weekdays(x))
  }
my_history$day <- sapply(my_history$Date, day_of_week)</pre>
sibling_history$day <- sapply(sibling_history$Date, day_of_week)
#Compare dataset
my_history$individual <- "Tay"</pre>
sibling_history$individual <- "Sis"</pre>
merged <- rbind(my_history, sibling_history)</pre>
#Numerical analysis
#Graphical Analysis
```

```
ggplot(merged, aes(x = day, fill=individual) ) +
  geom_bar(position = "dodge") +
  labs("Sis v. Tay: Netflix shows watched by day")
```





April-2020	May-2020 June-2020	100
28 29 30 31 1 2 3	25 26 27 28 29 30 1 30 31 1 2 3 4 5	
4 5 6 7 8 9 10	2 3 4 5 6 7 8 6 7 8 9 10 11 12	
11 12 13 14 15 16 17	9 10 11 12 13 14 15 13 14 15 16 17 18 19	
18 19 20 21 22 23 24	16 17 18 19 20 21 22 20 21 22 23 24 25 26	
25 <mark>26 27 28 29 30</mark> 1	23 24 25 26 27 28 29 27 28 29 30 1 2 3	80
2 3 4 5 6 7 8	30 31 1 2 3 4 5 4 5 6 7 8 9 10	
SSMTWTF	SSMTWTF SSMTWTF	
July-2020	August-2020 eptember-2020	
27 28 29 30 1 2 3	25 26 27 28 29 30 31 29 30 31 1 2 3 4	60
4 5 6 7 8 9 10	1 2 3 4 5 6 7 5 6 7 8 9 10 11	
11 12 13 14 15 16 17	8 9 10 11 12 <mark>13 14 12 13 14 15 16 17 18</mark>	
18 19 20 21 22 23 24	15 16 17 18 19 20 21 19 20 21 22 23 24 25	
25 26 27 28 29 30 31	22 23 24 25 <mark>26 27</mark> 28	
1 2 3 4 5 6 7	29 30 31 1 2 3 4 3 4 5 6 7 8 9	40
SSMTWTF	SSMTWTF SSMTWTF	
October-2020	November-2020 December-2020	
26 27 28 29 30 1 2	31 1 2 3 4 5 6 28 29 30 1 2 3 4	
3 4 5 6 7 8 9	7 8 9 10 11 12 13 5 6 7 8 9 10 11	20
10 11 12 13 14 15 16	14 15 16 17 18 <mark>19 20 </mark>	
17 18 19 20 21 22 23	21 22 23 24 25 26 27 19 20 21 22 23 24 25	
24 25 26 27 28 29 30	28 29 30 1 2 3 4 26 27 28 29 30 31 1	
31 1 2 3 4 5 6	5 6 7 8 9 10 11 2 3 4 5 6 7 8	
SSMTWTF	SSMTWTF SSMTWTF	